DOCUMENT RESUME

ED 427 020 TM 029 360

AUTHOR Newman, Isadore; Fraas, John W.
TITLE The Responsibility of Educational Researchers To Make Appropriate Decisions about the Error Rate Unit on Which Type I Error Adjustments Are Based: A Thoughtful Process Not a Mechanical One.
PUB DATE 1998-10-00
NOTE 21p.; Paper presented at the Annual Meeting of the Midwestern Educational Research Association (Chicago, IL, October 1998).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Decision Making; *Educational Research; *Error of Measurement; Program Evaluation; *Researchers; *Responsibility
IDENTIFIERS *Type I Errors

ABSTRACT

Educational researchers often use multiple statistical tests in their research studies and program evaluations. When multiple statistical tests are conducted, the chance that Type I errors may be committed increases. Thus, the researchers are faced with the task of adjusting the alpha levels for their individual statistical tests in order to keep the overall alpha value at a reasonable level. A three-step procedure is presented that can be used to adjust the alpha levels of the individual statistical tests. This procedure requires researchers to: (1) identify the appropriate conceptual unit for the error rate (pairwise, experimentwise, and familywise); (2) determine the number and nature of tests contained in that error rate unit; and (3) apply a Bonferroni-type adjustment procedure to the various statistical tests contained in each error rate unit. This three-step adjustment procedure emphasizes that it is the obligation of the researchers to make logical decisions, not mechanical ones, when adjusting Type I error rates for multiple statistical tests. (Contains 20 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

Running head: The Responsibility of Educational Researchers

ED 427 020

The Responsibility of Educational Researchers to Make Appropriate Decisions About The Error Rate Unit on Which Type I Error Adjustments Are Based: A Thoughtful Process Not a Mechanical One

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

John Fraas

Isadore Newman

The University of Akron

John W. Fraas

Ashland University

Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, October 1998, Chicago, Illinois

TM029360

BEST COPY AVAILABLE

Abstract

Educational researchers often use multiple statistical tests in their research studies and program evaluations. When multiple statistical tests are conducted, the chance that Type I errors may be committed increases. Thus, the researchers are faced with the task of adjusting the alpha levels for their individual statistical tests in order to keep the overall alpha value at a reasonable level. In this paper we present a three-step procedure that can be used to adjust the alpha levels of the individual statistical tests. This procedure requires researchers to: (a) identify the appropriate conceptual unit for the error rate (pairwise, experimentwise, and familywise), (b) determine the number and nature of the tests contained in that error rate unit, and (c) apply a Bonferroni-type adjustment procedure to the various statistical tests contained in each error rate unit. This three-step adjustment procedure emphasizes that it is the obligation of the researchers to make logical decisions, not mechanical ones when adjusting Type I error rates for multiple statistical tests.

The Responsibility of Educational Researchers to Make Appropriate Decisions About The Error Rate Unit on Which Type I Error Adjustments Are Based

A Thoughtful Process not a Mechanical One

As noted by Stevens (1996), it is common for multiple statistical tests to be conducted in educational studies and program evaluations. Stevens suggests that multiple statistical tests are encountered in studies that researchers may not readily recognize as the type of studies that contain multiple statistical tests. These types of studies include (a) ANOVA designs that contain main effects and interaction effects, (b) analyses of multiple dependent variables, (c) analyses of multiple regression models, and (d) analyses of numerous correlation coefficients.

The use of multiple statistical tests may lead to a situation where the chance of committing at least one Type I error, i.e., a true null hypothesis will be rejected, is significantly increased. To understand the impact of multiple statistical tests on the probability of committing at least one Type I error, assume that the researchers identified the error unit as the error rate per comparison, that is, each statistical test is considered as its own unit of error. If a statistical test is conducted with a selected alpha level of .05, the probability of committing a Type I error is .05 for that test. For two statistical tests with the alpha level set at .05 for each test, however, the probability of at least one Type I error approaches .10. The probability of committing at least one Type I error for m orthogonal statistical tests with the alpha value for a individual test set at α_{ind} can be calculated as follows:

$$p(\text{at least one Type I error}) = 1 - (1 - \alpha_{ind})^{m} \leq m\,\alpha_{ind} \qquad (1.1)$$

where $m\alpha_{ind}$ is the approximate upper bound on the probability of committing at least one Type I error. If the m statistical tests are not orthogonal, the probability of committing at least one Type I error can be summarized as follows:

$$p(\text{at least one Type I error}) \leq 1 - (1 - \alpha_{ind})^{m} \leq m\,\alpha_{ind} \qquad (1.2)$$

The empirical research indicates the probability of committing at least one Type I error is fairly close to $1 - (1 - \alpha_{ind})^{m}$ for statistical tests that are not orthogonal (Toothaker, 1991).

As revealed by Equations 1.1 and 1.2 and as noted by Kirk (1982), Stevens (1996), Hays (1988), Winer, Brown, and Michels (1991), and Toothaker (1991), the chance of committing a Type I error can increase dramatically when multiple statistical tests are conducted. The previously mentioned authors provide detailed presentations of numerous methods that can be used to adjust the Type I error rates for multiple statistical tests. These methods include (a) Fisher's least significance differences test (1949), (b) Tukey's HSD test (1953), (c) Spjotvoll and Stoline's modification of the HSD test (1973), (d) Tukey-Kramer modification of the HSD test (Tukey, 1953; Kramer, 1956), (e) Scheffe's test (1953), (f) Brown-Forsythe BF procedure (1974), (g) Newman-Keuls test (Newman, 1939; Keuls, 1952), (h) Duncan's new multiple range test (Duncan, 1955), and (i) Bonferroni's procedure (Dunn, 1961; Newman & Fry, 1972).

The selection of an appropriate adjustment technique can be a daunting task for even an experienced researcher or program evaluator. More importantly, in today's advanced computer age, it is not uncommon for researchers and program evaluators to mechanically select a Type I error adjustment procedure with little thought being given to how and why the adjustment is being made.

BEST COPY AVAILABLE

The purpose of this paper is to present a three-step adjustment approach that can easily be applied by researchers and program evaluators to numerous types of multiple statistical testing situations. This approach is very generalizable, i.e., it can be applied to studies in which one wishes to control for various types of comparisons including pairwise, familywise, and experimentwise. In addition, it can be used with studies that contain repeated measures, covariance analyses, and/or unequal sample sizes.

We believe that this adjustment procedure has an even more important characteristic than those previously mentioned. This three-step procedure requires the researchers to address two questions:

1. What is the appropriate conceptual error rate unit?

2. What specific adjustment should be made for each statistical test in that unit in order to maintain an acceptable overall alpha level?

The act of addressing these two questions prevents researchers from mechanically implementing Type I error adjustments to their multiple statistical tests. Rather, it requires thoughtfulness on the part of the researchers. Thus, the use of this procedure should provide researchers and program evaluators with a better understanding of their analyses.

Hypothetical Study

To illustrate how the three-step procedure can be used by researchers and program evaluators, we will refer to a hypothetical study that involves three groups of public school teachers. The study is designed to evaluate the impact that three instructional methods have on the participants' teaching efficacy levels. In this study, two elements of teacher efficacy that will be recorded for each participant are personal efficacy and teaching efficacy. The personal efficacy

scores will measure the degree that a teacher believes she he can personally impact student learning. The teaching efficacy scores will measure the degree that a teacher believes teachers, in general, can impact student learning. The participants' personal and teaching efficacy levels will be measured before and after they were exposed to one of the three instructional methods. The post-treatment personal efficacy scores and the post-treatment teaching efficacy scores will serve as the dependent variables.

Various hypotheses related to the participants' post-treatment personal and teaching efficacy scores are posed in this study. Once the data are collected, the researchers will statistically test the hypotheses through the use of multiple regression models. The independent variables included in the various regression models represent: (a) the teachers' pre-treatment personal and teaching efficacy scores, (b) a series of dummy variables that identify the three treatment groups, and (c) the product of each of the pre-treatment score variables and each of the treatment dummy variables, which are required to test the interaction effects. See Table 1 for a list of variables used in the various regression models.

Insert Table 1 about here

In this study, the researchers are interested in determining whether the data support the following research hypotheses:

$1H_1$: An interaction effect between the pre-treatment personal efficacy scores and the treatments account for some of the variation in the post-treatment personal efficacy scores.

$2H_1$: An interaction effect between the pre-treatment teaching efficacy scores and the treatments account for some of the variation in the post-treatment teaching efficacy scores.

The full and restricted regression models that will be used to determine if $1H_1$ is supported by the data are as follows:

$$Y_1 = a_1U + b_1X_1 + b_2X_2 + b_3X_3 + b_5X_5 + b_6X_6 + e_1 \quad \text{[Model 1]}$$

$$Y_1 = a_2U + b_1X_1 + b_2X_2 + b_4X_4 + e_2 \quad \text{[Model 2]}$$

The probability produced by the F test of the difference between the R^2 values of Model 1 and Model 2 will be used to determine if $1H_1$ is supported by the data. That is, this probability value will be compared to the alpha level set by the researchers.

The full and restricted regression models that will be used to determining whether $2H_1$ is supported by the data are as follows:

$$Y_2 = a_3U + b_1X_1 + b_2X_2 + b_3X_3 + b_{10}X_{10} + b_{11}X_{11} + e_3 \quad \text{[Model 3]}$$

$$Y_2 = a_4U + b_1X_1 + b_2X_2 + b_5X_5 + e_4 \quad \text{[Model 4]}$$

The probability of the F test of the difference between the R^2 values of Model 3 and Model 4 will be used to determine if $2H_1$ is supported by the data. Again, this probability value will be compared to the alpha level set by the researchers.

If the researchers find that the data do not support $1H_1$, they will determine whether the data support the following additional hypotheses:

$3H_1$: At least one difference exists among the post-treatment personal efficacy scores of the three treatments adjusting for the pre-treatment personal efficacy scores.

$4H_1$: At least one difference exists among the post-treatment teaching efficacy scores of the three treatments adjusting for the pre-treatment teaching efficacy scores.

The full and restricted regression models that will be used to determine whether $3H_1$ is supported by the data are as follows:

$$Y_1 = a_2U + b_1X_1 + b_2X_2 + b_4X_4 + e_2 \quad \text{[Model 2]}$$

$$Y_1 = a_5U + b_4X_4 + e_5 \quad \text{[Model 5]}$$

The probability corresponding to the F test of the difference between the R^2 values of Model 2 and Model 5 will be used to determine whether the data support $3H_1$. This probability value will be compared to the alpha level set by the researchers.

If $3H_1$ is supported by the data, the researchers will determine which adjusted treatment means differ. That is, the three pairwise comparisons of the adjusted post-treatment personal efficacy means will be conducted. The following regression models will be used to determine whether differences exist between the adjusted post-treatment personal efficacy means of Treatment 1 versus Treatment 2, Treatment 1 versus Treatment 3, and Treatment 2 versus Treatment 3.

$$Y_1 = a_2U + b_1X_1 + b_2X_2 + b_4X_4 + e_2 \quad \text{[Model 2]}$$

$$Y_1 = a_6U + b_2X_2 + b_3X_3 + b_4X_4 + e_6 \quad \text{[Model 6]}$$

The probability of the t tests of the b_1 and b_2 coefficients in Model 2 will be compared to the established alpha levels to determine if differences exist between the adjusted post-treatment personal efficacy means of Treatment 1 versus Treatment 3 and Treatment 2 versus Treatment 3. The probability of the t test of the b_2 coefficient in Model 6 will be compared to the established

alpha level to determine if a difference exists between the adjusted post-treatment personal efficacy means of Treatment 2 versus Treatment 1.

In a similar fashion, if $4H_1$ is supported by the data, the researchers will determine which adjusted post-treatment teaching efficacy means of the treatments differ. The following regression models will be used to determine whether differences exist between the adjusted post-treatment teaching efficacy means of Treatment 1 versus Treatment 2, Treatment 1 versus Treatment 3, and Treatment 2 versus Treatment 3.

$$Y_2 = a_4U + b_1X_1 + b_2X_2 + b_3X_3 + e_4 \quad \text{[Model 4]}$$

$$Y_2 = a_7U + b_2X_2 + b_4X_4 + b_5X_5 + e_7 \quad \text{[Model 7]}$$

The probability of the t tests of the b_1 and b_2 coefficients in Model 4 will be compared to the established alpha levels to determine if differences exist between the adjusted post-treatment teaching efficacy means of Treatment 1 versus Treatment 3 and Treatment 2 versus Treatment 3. The probability of the t test of the b_2 coefficient in Model 7 will be compared to the established alpha level to determine if a difference exists between the adjusted post-treatment teaching efficacy means of Treatment 2 versus Treatment 1.

Before statistical testing is conducted in this study, it is important for the researchers to identify the number and nature of the statistical tests that will be tested. The 10 statistical tests that the hypothetical study contains include the following:

1. A maximum of five statistical tests could be conducted on the personal efficacy scores. A statistical test will be used to test for the existence of an interaction effect between the pre-treatment personal efficacy scores and the treatments. If this interaction effect is not significant, another statistical test will be used to test for at least one difference among the adjusted post-

treatment personal efficacy means of the three treatments. If the test of the adjusted means suggests that at least two adjusted means differ among the three treatments, three statistical tests of the pairwise comparisons will conducted.

2. A maximum of five corresponding statistical tests could be conducted on the post-treatment teaching efficacy scores.

The issue facing the researchers is: What alpha levels should be used for the various statistical tests? The following section of this paper illustrates how a three-step adjustment procedure can be used to determine the alpha levels for the individual statistical tests.

A Three-Step Type I Error Adjustment Procedure

We take the position that controlling the Type I error rates in studies that contain multiple statistical tests requires the researchers to address three major issues. First, the researcher must identify the appropriate conceptual rate error units for which the adjustments will be made. Second, once this identification process has been completed, the researchers must determine the number and nature of statistical tests included in each error rate unit. Third, the researchers must implement a technique that will adjust the alpha level of each statistical test, if necessary, contained in the error rate unit. Although there are numerous adjustment methods that could be used, we strongly recommend that researchers implement a Bonferroni-type adjustment procedure. In spite of the fact that a Bonferroni-type adjustment procedure tends to be more conservative, that is, it has less power, than some other types of procedures in certain situations, we believe that such an adjustment procedure gives researchers greater flexibility when dealing with multiple statistical analyses and complex research designs.

BEST COPY AVAILABLE

Step One: Defining the Error Rate Units

In order to adjust the Type I error rates for a study that involves multiple statistical tests, the researchers must first specify the appropriate conceptual unit for the error rate (Kirk, 1982, Hays, 1988; Winer, Brown, & Michels, 1991; Toothaker, 1991). As noted by a number of authors, the relative merits of using one conceptual error unit over another can be debated (Duncan, 1955; McHugh & Ellis, 1955; Ryan, 1959, 1962; Wilson, 1962). Nevertheless, the identification of a conceptual error rate unit, which we find is seldom discussed in research studies that include multiple statistical tests, is an important element in the evaluation of a study's results. The selection of the error rate unit forms the logical framework on which the adjustments of the alpha levels of the individual statistical tests are based. Researchers should be able to clearly identify and justify the formation of the different error rate units.

In some studies, researchers may be faced with various choices regarding how to define the error rate unit. To illustrate this point, again consider the teacher efficacy study. The researchers could consider all 10 of the statistical tests as one error rate unit. On the other hand, they may identify two error rate units with each unit based on one of the two dependent variables. Thus, two possible delineations of error rate units for the hypothetical teaching efficacy study are as follows:

1. Delineation 1 -- The researchers form only one error rate unit for the entire study.

2. Delineation 2 -- The researchers identify two error rate units, which are based on the two dependent variables.

Some error rate units are referred to as experimentwise error units and others are referred to as familywise error units. Rather than being overly concerned with whether the unit is

experimentwise or familywise, we believe that it is more important that researchers and program evaluators clearly identity the error rate units that are being delineated and the number and nature of the statistical tests included in each of those units.

It should be noted that the researchers' choice of the error rate unit on which the adjustments are based could be a factor in determining whether the statistical test results reported in a study are statistically significant or not. Thus, analogous to the often suggested practice of reporting a R^2 value along with the corresponding test of significance as a means of facilitating the interpretation of statistical results, we strongly recommend that not only should the error rate units be clearly identified in the study, but the researchers should also present the logic on which these error rate units are based.

Step 2: Identify the Number and Nature of the Statistical Tests in Each Error Rate Unit.

As revealed by Equations 1.1 and 1.2, the probability of committing at least one Type I error for a given number of statistical tests contained in a specified error rate unit is determined by two factors: (a) the alpha level established for each statistical test (α_{mod}) and (b) the number of statistical tests (m). Before researchers can determine the appropriate alpha level for each statistical test, they must determine the number and nature of the statistical test contained in the error rate unit.

In the teacher efficacy study, the researchers would identify the number and nature of the statistical tests contained in the two different groupings of error rate units as follows:

1. Delineation 1 -- Since only one error rate unit is specified under this identification process, all 10 statistical tests included in this study are contained in this error rate unit. These

10 tests include two statistical tests of interaction effects, two statistical tests of the treatment main effects, and six statistical tests of the pairwise comparisons of the adjusted treatment means.

2. Delineation 2 -- Since this delineation contains two error rate units that are based on the two dependent variables, each of these two error rate units contains five statistical tests. The error rate unit based on the personal efficacy scores contains the statistical test of the interaction effect, the statistical test of the treatment main effect, and the three statistical tests of the pairwise comparisons of the adjusted treatment means. The second error rate unit, which is based on the teaching efficacy scores, contains five statistical tests that correspond to the tests contained in the first error rate unit.

Once the number and the nature of statistical tests contained in each error unit has been identified, the alpha level for each statistical test contained in a given error unit should be adjusted in order to prevent the overall alpha level of that error rate unit from becoming inflated to an unacceptable level. This task is addressed in the third and final step of this three-step adjustment procedure.

Step 3: Adjusting the Alpha Levels of the Various Statistical Tests.

We suggest researchers and program evaluators will find that an adjustment approach based on the Bonferroni inequality will provide them with a very practical and robust method of adjusting the alpha level for each statistical test in a given error rate unit. To illustrate how the adjustment procedure can be implemented, assume that the researchers want to maintain the alpha level near the .05 for each of the error rate units identified by the researchers in the teacher efficacy study. The alpha levels for the individual statistical tests would be calculated as follows for the two delineations of error rate units previously discussed.

1. Delineation 1 -- The researchers identified only one error rate unit in this approach and that unit contained two statistical tests of the interaction effects, two statistical tests of the treatment main effect, and six statistical tests for the pairwise comparisons of the adjusted treatment means. Thus, the alpha levels for the statistical tests of the two interaction effects and the two treatment main effects tests will be set at .0125, which is obtained by dividing .05 by four. If either of the statistical tests of the treatment main effects is significant, the alpha levels for each pairwise comparison test is set at .0021, which is calculated by dividing the adjusted alpha level used for both treatment main effect statistical tests (.0125) by the number of pairwise comparisons of the adjusted treatment means (6).

2. Delineation 2 -- The two error rate units identified in this approach are based on the two dependent variables, i.e., the two types of efficacy scores. Each error unit contains one statistical test of the interaction effect, one statistical test of the treatment main effect, and the statistical tests of the three pairwise comparisons of the adjusted treatment means. Thus, the alpha level used to statistically test the interaction effect and the treatment main effect for each dependent variable is .025, which is obtained by dividing the .05 alpha level by the total number of interaction effects (1) and treatment main effects (1) contained in the error rate unit. The alpha level for each pairwise comparison test of adjusted treatment means is set at .0083, which is obtained by dividing the adjusted alpha level for the treatment main effect (.025) by the number of pairwise comparisons of the adjusted treatment means contained in the error rate unit (3).

Once the alpha levels are adjusted, the researchers would simply compare the probability of a given statistical test to its adjusted alpha level to determine if the null hypothesis should be rejected. Implementing this three-step adjustment procedure not only provides protection against

inflated Type I error rates in studies that involve multiple statistical tests, but it will also forces researchers to reflect on the rationale that they use to make those adjustments

Two Additional Issues

It should be noted that two additional issues, which we did not specifically address in the examples discussed thus far in this paper, should be part of the reflection process that one uses to adjust Type I error rates. The first issue deals with the question: Are your statistical tests planned, i.e., do your statistical tests come from a strong theoretical, empirical and/or logical base? The second issue deals with the question: Are your statistical tests orthogonal? Consideration of these two questions by the researchers is also an essential element in the adjustment process of Type I error rates.

To illustrate this point with respect to planned tests, assume that we decided to form one error unit in the hypothetical teacher efficacy study, i.e., Delineation 1 is used. In addition, we have a theoretical reason for assuming that interaction effects will exist for both the personal and teaching efficacy scores. In this case, we may decide not to adjust the alpha levels for the two statistical tests of the interaction effects, but we would adjust the alpha levels for the subsequent statistical tests, which are the statistical tests of the treatment main effects and the pairwise comparisons of the adjusted treatment means. Thus, the alpha level for each interaction effect would be set at .05. The alpha levels for the two treatment main effects, however, would be set at .025, which is obtained by dividing .05 by the number of treatment effects (2). The alpha levels for each pairwise comparison of the adjusted treatment means would be set at .004, which is equal to the alpha level used for the group main effect (.025) divided by the number of pairwise comparisons being tested (6).

To further illustrate this point, assume we identify two error rate units such as those presented in Delineation 2. Also assume that empirical evidence exists that would allow us to expect that interaction effects will be present for both the teaching and personal efficacy scores. Since we expected statistically significant interaction effects, we would set the alpha level for each test of the interaction effects at .05. Due to the fact that each error rate unit contains only one treatment main effect and the interaction effect test contained in that error rate unit is not part of the adjustment process, the alpha level for each treatment main effect would be set at .05. In both error rate units the alpha level for each pairwise comparison test of the adjusted treatment means would be set at .017, which is obtained by dividing the alpha level of the treatment main effect (.05) by the number of pairwise comparisons (3) contained in each error rate unit.

Summary

To protect against inflated Type I error rates in studies that contain multiple statistical tests, researchers need to consider implementing an adjustment process. The three-step adjustment procedure presented in this paper, which is based upon a Bonferroni-type adjustment process, first requires the researchers to identify the error rate unit or units (pairwise, familywise, and experimentwise) that will serve as the basis for the adjustment process. Second, they must identify the number and nature of the statistical tests contained in each error rate unit. Finally, they must adjust the alpha levels of some, if not all, of the individual statistical tests contained in the error rate unit.

This three-step adjustment procedure provides researchers with a tool that is robust, flexible, and easy to apply. More importantly, however, this procedure requires researchers to

reflect on the selection of the conceptual error rate unit on which the Type I error adjustments are based. Such reflection should lead to better research and program evaluation.

In addition to encouraging the use of this three-step adjustment approach, we believe that it is important for educational researchers to engage in philosophical discussions and research to identify the most appropriate error rate units for specific types of research questions and situations. We believe that this type of investigation may prove to be just as valuable, if not more so, for the fields of educational research and program evaluation than additional Monte Carlo and analytical studies on Type I error rate correction procedures.

References

Brown, M B , & Forsythe, A B (1974) The ANOVA and multiple comparisons for data with heterogeneous variance Biometrics, 30, 719-724

Duncan, D B (1955) Multiple range and multiple F tests Biometrics, 11, 1-42

Dunn, O J (1961) Multiple comparisons among means Journal of the American Statistical Association, 56, 52-64

Fisher, R A (1949) The design of experiments Edinburgh Oliver and Boyd Ltd

Hays, W L (1988) Statistics (4th ed) Fort Worth, TX Holt, Rinehart and Winston

Kramer, C Y (1956) Extension of multiple range test to group means with unequal numbers of replications Biometrics, 12, 307-310

Kirk, R E (1982) Experimental design Procedures for the behavioral sciences (2nd ed) Belmont, CA Brooks/Cole

Kuels, M (1952) The use of studentized range in connection with an analysis of variance Euphytica, 1, 112-122

McHugh, R B , & Ellis, D S (1955) The "postmortem" testing of experimental comparisons Psychological Bulletin, 52, 425-428

Newman, D (1939) The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation Biometrika, 31, 20-30

Newman, I , & Fry, J (1972) A note on multiple comparisons and a comment on shrinkage Multiple Linear Regression Viewpoints, 2, (3), 36-39

Ryan, T A (1959) Multiple comparisons in psychological research Psychological Bulletin, 56, 26-47

Ryan, T A (1962) The experiment as the unit for computing rates of error Psychological Bulletin, 59, 301-305

Scheffé, H A method for judging all contrasts in the analysis of variance Biometrika, 40, 87-104

Spjøtvoll, E , & Stoline, M R An extension of the T-method of multiple comparisons to include the cases with unequal sample signs Journal of the American Statistical Association, 68, 975-978

Stevens, J (1996) Applied multivariate statistics for the social sciences (3rd ed) Mahwah, NJ Erlbaum

Toothaker, L E (1991) Multiple comparisons for researchers Newbury Park, CA Sage

Tukey, J W (1953) The problem of multiple comparisons Ditto Princeton University

Wilson, W A note on the inconsistency inherent in the necessity to perform multiple comparisons Psychological Bulletin, 59, 296-300

Winer, B J , Brown, D R ; & Michols, K M (1991) Statistical principles in experimental design (3rd ed) New York McGraw-Hill

Table 1

Description of the Variables and Symbols Used in the Regression Models

Symbol	Description
Y_1	Post-Treatment Personal Efficacy Scores
Y_2	Post-Treatment Teaching Efficacy Scores
X_1	Treatment 1 (If in Treatment 1, $X_1 = 1$, otherwise $X_1 = 0$)
X_2	Treatment 2 (If in Treatment 2, $X_2 = 1$, otherwise $X_2 = 0$)
X_3	Treatment 3 (If in Treatment 3, $X_3 = 1$, otherwise $X_3 = 0$)
X_4	Pre-Treatment Personal Efficacy Scores
X_5	Pre-Treatment Teaching Efficacy Scores
X_6	$X_1 * X_4$
X_7	$X_2 * X_4$
X_8	$X_3 * X_4$
X_9	$X_1 * X_5$
X_{10}	$X_2 * X_5$
X_{11}	$X_3 * X_5$
u	unit vector
a_{1-7}	constant term
e_{1-7}	error terms for the various regression models